



Numerical comparisons of two effective methods for mixed complementarity problems

Jein-Shan Chen^{a,*,1}, Shaohua Pan^b, Ching-Yu Yang^a

^a Department of Mathematics, National Taiwan Normal University, Taipei, Taiwan 11677, Taiwan

^b School of Mathematical Sciences, South China University of Technology, Guangzhou 510640, China

ARTICLE INFO

Article history:

Received 25 March 2009

Received in revised form 4 January 2010

Keywords:

MCP

The generalized FB function

Semismooth

Convergence rate

ABSTRACT

Recently there have two different effective methods proposed by Kanzow et al. in (Kanzow, 2001 [8]) and (Kanzow and Petra, 2004 [9]), respectively, which commonly use the Fischer–Burmeister (FB) function to recast the mixed complementarity problem (MCP) as a constrained minimization problem and a nonlinear system of equations, respectively. They all remark that their algorithms may be improved if the FB function is replaced by other NCP functions. Accordingly, in this paper, we employ the generalized Fischer–Burmeister (GFB) where the 2-norm in the FB function is relaxed to a general p -norm ($p > 1$) for the two methods and investigate how much the improvement is by changing the parameter p as well as which method is influenced more when we do so, by the performance profiles of iterations and function evaluations for the two methods with different p on MCPLIB collection.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The mixed complementarity problem (MCP) arises in many applications including the fields of economics, engineering, and operations research [1–4] and has attracted much attention in the past decade [5–10]. A collection of nonlinear mixed complementarity problems called MCPLIB can be found in [11] and an excellent book [12] is a good source for seeking theoretical backgrounds and numerical methods for it.

Given a mapping $F : [l, u] \rightarrow \mathbb{R}^n$ with $F = (F_1, \dots, F_n)^T$, where $l = (l_1, \dots, l_n)^T$ and $u = (u_1, \dots, u_n)^T$ with $l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$ and $l_i < u_i$ for $i = 1, 2, \dots, n$. The MCP is to find a vector $x^* \in [l, u]$ such that each component x_i^* satisfies exactly one of the following implications:

$$\begin{aligned} x_i^* = l_i &\implies F_i(x^*) \geq 0, \\ x_i^* \in (l_i, u_i) &\implies F_i(x^*) = 0, \\ x_i^* = u_i &\implies F_i(x^*) \leq 0. \end{aligned} \quad (1)$$

It is easy to see that, when $l_i = -\infty$ and $u_i = +\infty$ for all $i = 1, 2, \dots, n$, MCP (1) is equivalent to solving the nonlinear system of equations

$$F(x) = 0; \quad (2)$$

* Corresponding author. Tel.: +886 2 29325417; fax: +886 2 29332342.

E-mail addresses: jschen@math.ntnu.edu.tw (J.-S. Chen), shhpan@scut.edu.cn (S. Pan), yangcy@abel.math.ntnu.edu.tw (C.-Y. Yang).

¹ Member of Mathematics Division, National Center for Theoretical Sciences, Taipei Office, Taiwan.

when $l_i = 0$ and $u_i = +\infty$ for all $i = 1, 2, \dots, n$, it reduces to the nonlinear complementarity problems (NCP) which is to find a point $x \in \mathbb{R}^n$ such that

$$x \geq 0, \quad F(x) \geq 0, \quad \langle x, F(x) \rangle = 0. \quad (3)$$

In fact, from Theorem 2 of [13], MCP (1) itself is equivalent to the famous variational inequality problem (VIP) which is to find a vector $x^* \in [l, u]$ such that

$$\langle F(x^*), x - x^* \rangle \geq 0 \quad \forall x \in [l, u]. \quad (4)$$

Unless otherwise stated, the mapping F is assumed to be continuously differentiable.

Many methods have been proposed for the solution of MCP (1), among which there are two effective methods that attract much attention recently. They are the strictly feasible equation-based methods [6–8] and the semismooth Levenberg–Marquardt methods [9,10]. Some other variants of these methods can be found in [14–16]. The ideas for the aforementioned two methods are to reformulate (1) as a constrained minimization or a nonsmooth system of equations by using the Fischer–Burmeister function

$$\phi_{\text{FB}}(a, b) := \sqrt{a^2 + b^2} - (a + b) \quad \forall a, b \in \mathbb{R}. \quad (5)$$

The strictly feasible Newton-type method was considered in [8] to overcome drawbacks of some typical solution methods for the MCP (see e.g. [7]), for example, they can generate feasible iterates but have to solve relatively complicated subproblems or they have simple subproblems but do not necessarily generate feasible iterates. On the other hand, the semismooth Levenberg–Marquardt method was proposed in [9] to overcome some drawbacks of equation-based methods using the FB function. This method has the advantages that gradient steps are not necessary to obtain global convergence and it is more robust than those equation-based methods based on the FB function.

Recently, an extension of the FB function was considered in [17–19] by two of the authors. Specifically, they define the generalized Fischer–Burmeister (GFB) function by

$$\phi_p(a, b) := \|(a, b)\|_p - (a + b) \quad \forall a, b \in \mathbb{R}, \quad (6)$$

where p is an arbitrary fixed real number from the interval $(1, +\infty)$ and $\|(a, b)\|_p$ denotes the p -norm of (a, b) , i.e., $\|(a, b)\|_p = \sqrt[p]{|a|^p + |b|^p}$. In other words, in the function ϕ_p , they replace the 2-norm of (a, b) involved in the FB function by a more general p -norm. The function ϕ_p is still an NCP-function, that is, it satisfies the equivalence

$$\phi_p(a, b) = 0 \iff a \geq 0, \quad b \geq 0, \quad ab = 0. \quad (7)$$

For any given $p > 1$, the function ϕ_p was shown to possess all favorable properties of ϕ_{FB} ; see [17–19]. For example, its square is continuously differentiable everywhere on \mathbb{R}^2 .

In this paper, we follow the ideas used in the aforementioned two effective methods to solve MCP (1) whose solution may not be unique. For each method, we design a similar algorithm in which the GFB function is involved. We will present their convergence results although these results are analogous to those cases where ϕ_{FB} was considered. In fact, these convergence results are not hard to obtain since ϕ_{FB} and ϕ_p share almost the same favorable properties. However, the focus of this paper is on the numerical side as titled. We apply the two methods for solving all MCPLIB test problems, observe and analyze their numerical results. Furthermore, by the notion of performance profile introduced in [20], we plot the performances profile figures of iterations and function evaluations, respectively, for the two algorithms corresponding to four p . The performance profiles clearly and objectively reflect the influence of p on these two methods. Comparing Figs. 1–2 with Figs. 3–4, we see that the value of p has much more influence on the strictly feasible semismooth algorithm than the semismooth Levenberg–Marquardt algorithm. A larger p (for example over 10^3) or a smaller p (for example in $(1, 1.001]$) will lead to worse performance of the strictly feasible semismooth algorithm; whereas a small p (for example $p = 1.001$) will bring good performance to the semismooth Levenberg–Marquardt algorithm.

Throughout this paper, \mathbb{R}^n denotes the space of n -dimensional real column vectors with the usual Euclidean product $\langle \cdot, \cdot \rangle$. For every differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\nabla f(x)$ denotes the gradient of f at x , and for every differentiable mapping F , $\nabla F(x)$ denotes the transposed Jacobian of F at x . For a vector $x \in \mathbb{R}^n$, the notation $[x]_+$ means the projection of x on $[l, u]$, whereas for a scalar s , $(s)_+$ means the projection of s on \mathbb{R}_+ , i.e., $(s)_+ = \max\{0, s\}$. We denote $\|x\|_p$ the p -norm of x and $\|x\|$ the Euclidean norm of x .

2. Preliminaries

In this section, we review some basic concepts that will be used in the subsequent analysis. First, we introduce the concept of generalized Jacobian of a mapping. Let $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a locally Lipschitz continuous mapping. Then, G is almost everywhere differentiable by Rademacher's Theorem (see [21]). In this case, the *generalized Jacobian* $\partial G(x)$ of G at x (in the Clarke sense) is defined as the convex hull of the B-subdifferential

$$\partial_B G(x) := \{V \in \mathbb{R}^{m \times n} \mid \exists \{x^k\} \subseteq D_G : \{x^k\} \rightarrow x \text{ and } G'(x^k) \rightarrow V\},$$

where D_G is the set of differentiable points of G . In other words, $\partial G(x) = \text{conv } \partial_B G(x)$. If $m = 1$, we call $\partial G(x)$ the *generalized gradient* of G at x . The calculation of $\partial G(x)$ is usually difficult in practice, and Qi [22] proposed the so-called *C-subdifferential* of G :

$$\partial_C G(x)^T := \partial_{G_1}(x) \times \cdots \times \partial_{G_m}(x) \quad (8)$$

which is easier to compute than the generalized Jacobian $\partial G(x)$. Here, the right-hand side of (8) denotes the set of matrices in $\mathbb{R}^{n \times m}$ whose i th column is given by the generalized gradient of the i th component function G_i . By Proposition 2.6.2 of [21],

$$\partial G(x)^T \subseteq \partial_C G(x)^T. \quad (9)$$

We next introduce the definition of (strongly) semismooth function. The semismooth property is very important from computational point of view. In particular, it plays a fundamental role in the superlinear convergence analysis of generalized Newton methods [23–25]. Assume that $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is locally Lipschitz continuous. G is called *semismooth* at x if G is directionally differentiable at x and for any $V \in \partial G(x+h)$ and $h \rightarrow 0$,

$$G(x+h) - G(x) - Vh = o(\|h\|); \quad (10)$$

G is called *strongly semismooth* at x if G is semismooth at x and for any $V \in \partial G(x+h)$ and $h \rightarrow 0$,

$$G(x+h) - G(x) - Vh = O(\|h\|^2); \quad (11)$$

G is called a (strongly) semismooth function if it is (strongly) semismooth everywhere.

The following lemma lists some properties of ϕ_p whose proofs can be found in [17–19]. Such results are ground stones for getting the properties of Φ_p and $\tilde{\Phi}_p$ in what follows.

Lemma 2.1. Let $\phi_p : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be defined as in (6). Then, the following results hold.

- (a) ϕ_p is a strongly semismooth NCP-function.
- (b) ϕ_p is Lipschitz continuous with the Lipschitz constant L given by $L = \sqrt{2} + 2^{(1/p-1/2)}$ when $1 < p < 2$ and $L = 1 + \sqrt{2}$ when $p \geq 2$.
- (c) Given any point $(a, b) \in \mathbb{R}^2$, each element in the generalized gradient $\partial \phi_p(a, b)$ has the representation $(\xi - 1, \zeta - 1)$ where, if $(a, b) \neq (0, 0)$,

$$\xi = \frac{\text{sign}(a) \cdot |a|^{p-1}}{\|(a, b)\|_p^{p-1}} \quad \text{and} \quad \zeta = \frac{\text{sign}(b) \cdot |b|^{p-1}}{\|(a, b)\|_p^{p-1}}$$

- with $\text{sign}(\cdot)$ denotes the sign function, and otherwise $(\xi, \zeta) \in \mathbb{R}^2$ denotes an arbitrary vector satisfying $|\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1$.
- (d) For any $a, b \in \mathbb{R}$ and $p > 1$, there holds that

$$(2 - 2^{\frac{1}{p}}) |\min\{a, b\}| \leq |\phi_p(a, b)| \leq (2 + 2^{\frac{1}{p}}) |\min\{a, b\}|. \quad (12)$$

- (e) The square of ϕ_p is a continuously differentiable NCP function.

The following lemma establishes another property of ϕ_p , which plays a key role in the nonsmooth system reformulation of MCP (1) with the generalized FB function.

Lemma 2.2. Let $\phi_p : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be defined by (6). Then, the following limits hold.

- (a) $\lim_{l_i \rightarrow -\infty} \phi_p(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) = -\phi_p(u_i - x_i, -F_i(x))$.
- (b) $\lim_{u_i \rightarrow \infty} \phi_p(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) = \phi_p(x_i - l_i, F_i(x))$.
- (c) $\lim_{l_i \rightarrow -\infty} \lim_{u_i \rightarrow \infty} \phi_p(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) = -F_i(x)$.

Proof. Let $\{a^k\} \subseteq \mathbb{R}$ be any sequence converging to $+\infty$ as $k \rightarrow \infty$ and $b \in \mathbb{R}$ be any fixed number. We will prove $\lim_{k \rightarrow \infty} \phi_p(a^k, b) = -b$, and part (a) then follows by continuity arguments. Without loss of generality, assume that $a^k > 0$ for each k . Then,

$$\begin{aligned} \phi_p(a^k, b) &= a^k \left(1 + (|b|/a^k)^p\right)^{1/p} - a^k - b \\ &= a^k \left[1 + \frac{1}{p} \left(\frac{|b|}{a^k}\right)^p + \frac{1-p}{2p^2} \left(\frac{|b|}{a^k}\right)^{2p} + \cdots + \frac{(1-p) \cdots (1-pn+p)}{n!p^n} \left(\frac{|b|}{a^k}\right)^{np} \right. \\ &\quad \left. + o\left(\left(\frac{|b|}{a^k}\right)^{pn}\right)\right] - a^k - b \\ &= \frac{1}{p} \frac{|b|^p}{(a^k)^{p-1}} + \frac{1-p}{2p^2} \frac{|b|^{2p}}{(a^k)^{2p-1}} + \cdots + \frac{(1-p) \cdots (1-pn+p)}{n!p^n} \frac{|b|^{np}}{(a^k)^{np-1}} + \frac{(a^k)|b|^{np}}{(a^k)^{np}} o\left(\frac{|b|/a^k}{(|b|/a^k)^{pn}}\right) - b \end{aligned}$$

where the third equality is using the Taylor expansion of the function $(1+t)^{1/p}$ and the notation $o(t)$ means $\lim_{t \rightarrow 0} o(t)/t = 0$. Since $a^k \rightarrow +\infty$ as $k \rightarrow \infty$, we have $\frac{|b|^{np}}{(a^k)^{np-1}} \rightarrow 0$ for all n . This together with the last equation implies $\lim_{k \rightarrow \infty} \phi_p(a^k, b) = -b$. This proves part (a). Parts (b) and (c) are direct by part (a) and the continuity of ϕ_p . \square

To conclude this section, we present a lemma which will be used in the subsequent analysis.

Lemma 2.3 ([7, Proposition 6]). For all negative definite diagonal matrices $D_a, D_b \in \mathbb{R}^{n \times n}$, a matrix of the form $D_a + D_b M$ is nonsingular if and only if $M \in \mathbb{R}^{n \times n}$ is a P_0 -matrix.

3. Strictly feasible Newton-type method

For convenience, in the rest of this paper, we adopt the following notations of index sets:

$$\begin{aligned} I_l &:= \{i \in \{1, 2, \dots, n\} \mid -\infty < l_i < u_i = +\infty\}, \\ I_u &:= \{i \in \{1, 2, \dots, n\} \mid -\infty = l_i < u_i < +\infty\}, \\ I_{lu} &:= \{i \in \{1, 2, \dots, n\} \mid -\infty < l_i < u_i < +\infty\}, \\ I_f &:= \{i \in \{1, 2, \dots, n\} \mid -\infty = l_i < u_i = +\infty\}. \end{aligned} \quad (13)$$

With the generalized FB function, we define an operator $\Phi_p : \mathbb{R}^n \rightarrow \mathbb{R}^n$ componentwise as

$$\Phi_{p,i}(x) := \begin{cases} \phi_p(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ -\phi_p(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ \phi_p(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) & \text{if } i \in I_{lu}, \\ -F_i(x) & \text{if } i \in I_f, \end{cases} \quad (14)$$

where the minus sign for $i \in I_u$ and $i \in I_f$ is motivated by Lemma 2.2. In fact, all results of this paper would be true without the minus sign. Using the equivalence (7), it is easily verified that a vector $x^* \in \mathbb{R}^n$ solves (1) if and only if x^* is a solution of the nonlinear system of equations $\Phi_p(x) = 0$. This means that the squared norm of Φ_p induces a family of merit functions for (1) in the sense that the solution of (1) is equivalent to finding a minimizer of the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} \Psi_p(x) := \frac{1}{2} \|\Phi_p(x)\|^2, \quad (15)$$

with the corresponding objective value equal to 0. In this section, we study the strictly feasible Newton-type method based on the constrained nonlinear system of equations

$$\Phi_p(x) = 0, \quad x \in [l, u], \quad (16)$$

and globalized by the projected gradient-type method for the constrained minimization

$$\min_{x \in [l, u]} \Psi_p(x). \quad (17)$$

Before describing the specific iterative schemes, we present a few nice properties of the mapping Φ_p and the merit function Ψ_p that will be used in the subsequent analysis.

3.1. Properties of Φ_p and Ψ_p

The following proposition states the smoothness of Ψ_p and the semismoothness of Φ_p , which are direct by Lemma 2.1(a) and (e), and Theorem 19 of [26].

Proposition 3.1. Let Φ_p and Ψ_p be defined as in (14) and (15), respectively. Then,

- (a) the mapping Φ_p is semismooth, and moreover, it is strongly semismooth if F' is locally Lipschitz continuous.
- (b) The function Ψ_p is continuously differentiable everywhere.

The following technical lemma gives an expression for each element in the generalized Jacobian of Φ_p at any point x which plays an important role in the subsequent analysis.

Lemma 3.1. For any given $x \in \mathbb{R}^n$, we have $\partial \Phi_p(x)^T \subseteq \{D_a(x) + \nabla F(x) D_b(x)\}$, where $D_a(x), D_b(x) \in \mathbb{R}^{n \times n}$ are diagonal matrices whose diagonal elements are defined below:

- (a) For $i \in I_l$, if $(x_i - l_i, F_i(x)) \neq (0, 0)$, then

$$(D_a)_{ii}(x) = \frac{\text{sign}(x_i - l_i) |x_i - l_i|^{p-1}}{\|(x_i - l_i, F_i(x))\|_p^{p-1}} - 1, \quad (D_b)_{ii}(x) = \frac{\text{sign}(F_i(x)) |F_i(x)|^{p-1}}{\|(x_i - l_i, F_i(x))\|_p^{p-1}} - 1, \quad (18)$$

and otherwise

$$((D_a)_{ii}(x), (D_b)_{ii}(x)) \in \left\{ (\xi - 1, \zeta - 1) \in \mathbb{R}^2 \mid |\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1 \right\}. \quad (19)$$

(b) For $i \in I_u$, if $(u_i - x_i, -F_i(x)) \neq (0, 0)$, then

$$(D_a)_{ii}(x) = \frac{\text{sign}(u_i - x_i) |u_i - x_i|^{p-1}}{\|(u_i - x_i, -F_i(x))\|_p^{p-1}} - 1, \quad (D_b)_{ii}(x) = \frac{-\text{sign}(F_i(x)) |F_i(x)|^{p-1}}{\|(u_i - x_i, -F_i(x))\|_p^{p-1}} - 1,$$

and otherwise $((D_a)_{ii}(x), (D_b)_{ii}(x)) \in \left\{ (\xi - 1, \zeta - 1) \in \mathbb{R}^2 \mid |\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1 \right\}$.

(c) For $i \in I_{lu}$, $(D_a)_{ii}(x) = a_i(x) + b_i(x)c_i(x)$ and $(D_b)_{ii}(x) = b_i(x)d_i(x)$, where, if $(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) \neq (0, 0)$, then

$$a_i(x) = \frac{\text{sign}(x_i - l_i) \cdot |x_i - l_i|^{p-1}}{\|(x_i - l_i, \phi_p(u_i - x_i, -F_i(x)))\|_p^{p-1}} - 1,$$

$$b_i(x) = \frac{\text{sign}(\phi_p(u_i - x_i, -F_i(x))) \cdot |\phi_p(u_i - x_i, -F_i(x))|^{p-1}}{\|(x_i - l_i, \phi_p(u_i - x_i, -F_i(x)))\|_p^{p-1}} - 1,$$

and otherwise $(a_i(x), b_i(x)) \in \left\{ (\xi - 1, \zeta - 1) \in \mathbb{R}^2 \mid |\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1 \right\}$; and if $(u_i - x_i, -F_i(x)) \neq (0, 0)$, then

$$c_i(x) = \frac{-\text{sign}(u_i - x_i) \cdot |u_i - x_i|^{p-1}}{\|(u_i - x_i, -F_i(x))\|_p^{p-1}} + 1,$$

$$d_i(x) = \frac{-\text{sign}(F_i(x)) \cdot |F_i(x)|^{p-1}}{\|(u_i - x_i, -F_i(x))\|_p^{p-1}} + 1,$$

and otherwise $(c_i(x), d_i(x)) \in \left\{ (\xi + 1, \zeta + 1) \in \mathbb{R}^2 \mid |\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1 \right\}$.

(d) For $i \in I_f$, $(D_a)_{ii}(x) = 0$ and $(D_b)_{ii}(x) = -1$.

Proof. Let $\Phi_p(x) := [(\Phi_p)_1(x), (\Phi_p)_2(x), \dots, (\Phi_p)_n(x)]^T$. Then, from (8) and (9),

$$\partial \Phi_p(x)^T \subseteq \partial(\Phi_p)_1(x) \times \partial(\Phi_p)_2(x) \times \dots \times \partial(\Phi_p)_n(x) \quad (20)$$

where the latter denotes the set of all matrices whose i th row belongs to $\partial(\Phi_p)_i(x)$ for each i . With this in mind, we proceed to prove the lemma.

(a) For $i \in I_l$, we have $(\Phi_p)_i(x) = \phi_p(x_i - l_i, F_i(x))$. If $(x_i - l_i, F_i(x)) \neq (0, 0)$, then ϕ_p is continuously differentiable at such point, and moreover, by Lemma 2.1(c),

$$\nabla_a \phi_p(x_i - l_i, F_i(x)) = (D_a)_{ii}(x), \quad \nabla_b \phi_p(x_i - l_i, F_i(x)) = (D_b)_{ii}(x)$$

with $(D_a)_{ii}(x)$ and $(D_b)_{ii}(x)$ given by (18). Direct calculation with chain rule gives

$$\partial(\Phi_p)_i(x)^T = \{(D_a)_{ii}(x)e^i + \nabla F_i(x)(D_b)_{ii}(x)\}$$

where $e^i \in \mathbb{R}^n$ denotes the column vector whose i th element is 1 but zero elsewhere. If $(x_i - l_i, F_i(x)) = (0, 0)$, then using the generalized chain rule [21, Theorem 2.3.10] yields

$$\partial(\Phi_p)_i(x)^T \subseteq \{(D_a)_{ii}(x)e^i + \nabla F_i(x)(D_b)_{ii}(x)\},$$

where $(D_a)_{ii}(x)$ and $(D_b)_{ii}(x)$ are given by (19). Thus, we prove part (a).

(b) Since $(\Phi_p)_i(x) = -\phi_p(u_i - x_i, -F_i(x))$ for $i \in I_u$, following the same arguments as in part (a) gives the desired results.

(c) For $i \in I_{lu}$, $(\Phi_p)_i(x) = \phi_p(x_i - l_i, \phi_p(u_i - x_i, -F_i(x)))$. We denote

$$g_i(x) := \phi_p(u_i - x_i, -F_i(x)) \quad \text{and} \quad h_i(x) := (x_i - l_i, g_i(x)).$$

In other words, $(\Phi_p)_i(x) = \phi_p(h_i(x))$. We first argue that $\partial(\Phi_p)_i(x) = \partial\phi_p(h_i(x))\partial h_i(x)$.

If $(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) \neq (0, 0)$, i.e., $h_i(x) \neq (0, 0)$, clearly, ϕ_p is continuously differentiable at $h_i(x)$. In addition, the continuous differentiability of F along with the Lipschitz continuity of ϕ_p (by Lemma 2.1(b)) implies that h_i is locally Lipschitz. By [21, Theorem 2.6.6], we then have $\partial(\Phi_p)_i(x) = \partial\phi_p(h_i(x))\partial h_i(x)$.

If $(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) = (0, 0)$, i.e., $h_i(x) = (x_i - l_i, g_i(x)) = (0, 0)$, then ϕ_p is continuously differentiable at $(u_i - x_i, -F_i(x))$ since $u_i - x_i = u_i - l_i > 0$. Hence, $h_i(x)$ is continuously differentiable at x , which by the corollary to [21, Proposition 2.2.1] implies that h_i is strictly differentiable at x . Furthermore, ϕ_p is Lipschitz and convex by [19, Proposition 3.1(b)]. This implies that ϕ_p is regular everywhere due to [21, Proposition 2.3.6(b)]. Then applying [21, Theorem 2.3.9(iii)] gives $\partial(\Phi_p)_i(x) = \partial\phi_p(h_i(x))\partial h_i(x)$.

Next we look into $\partial\phi_p(h_i(x))$ and $\partial h_i(x)$, and try to write them out. Let $(a_i(x), b_i(x)) \in \partial\phi_p(h_i(x))$. Since $h_i(x) = (x_i - l_i, g_i(x))$, we have $\partial h_i(x)^T = \{(e^i, \sigma^i) \mid \sigma^i \in \partial g_i(x)\}$ and

$$\partial(\Phi_p)_i(x) = \{a_i(x)(e^i)^T + b_i(x)(\sigma^i)^T \mid \sigma^i \in \partial g_i(x)\}. \quad (21)$$

The terms of $a_i(x)$ and $b_i(x)$ can be obtained by following the same calculation as done in part (a) where we only replace $F_i(x)$ by $g_i(x)$, which turns out

$$\begin{aligned} a_i(x) &= \frac{\text{sign}(x_i - l_i) \cdot |x_i - l_i|^{p-1}}{\|(x_i - l_i, g_i(x))\|_p^{p-1}} - 1 \\ &= \frac{\text{sign}(x_i - l_i) \cdot |x_i - l_i|^{p-1}}{\|(x_i - l_i, \phi_p(u_i - x_i, -F_i(x)))\|_p^{p-1}} - 1, \\ b_i(x) &= \frac{\text{sign}(g_i(x)) \cdot |g_i(x)|^{p-1}}{\|(x_i - l_i, g_i(x))\|_p^{p-1}} - 1 \\ &= \frac{\text{sign}(\phi_p(u_i - x_i, -F_i(x))) \cdot |\phi_p(u_i - x_i, -F_i(x))|^{p-1}}{\|(x_i - l_i, \phi_p(u_i - x_i, -F_i(x)))\|_p^{p-1}} - 1 \end{aligned}$$

for $(x_i - l_i, g_i(x)) = (x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) \neq (0, 0)$; and

$$(a_i(x), b_i(x)) \in \left\{ (\xi - 1, \zeta - 1) \in \mathbb{R}^2 \mid |\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1 \right\}$$

for $(x_i - l_i, g_i(x)) = (0, 0)$. Besides, by applying part (b) (with an additional minus sign), $\partial g_i(x) = \{c_i(x)e^i + d_i(x)\nabla F_i(x)\}$ where, if $(u_i - x_i, -F_i(x)) \neq (0, 0)$,

$$\begin{aligned} c_i(x) &= \frac{-\text{sign}(u_i - x_i) \cdot |u_i - x_i|^{p-1}}{\|(u_i - x_i, -F_i(x))\|_p^{p-1}} + 1, \\ d_i(x) &= \frac{-\text{sign}(F_i(x)) \cdot |F_i(x)|^{p-1}}{\|(u_i - x_i, -F_i(x))\|_p^{p-1}} + 1, \end{aligned}$$

and otherwise $(c_i(x), d_i(x)) \in \left\{ (\xi + 1, \zeta + 1) \in \mathbb{R}^2 \mid |\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1 \right\}$. Plugging all the above into (21), it then follows that

$$\begin{aligned} \partial(\Phi_p)_i(x) &= \{a_i(x)(e^i)^T + b_i(x)[c_i(x)(e^i)^T + d_i(x)F'_i(x)]\} \\ &= \{[a_i(x) + b_i(x)c_i(x)](e^i)^T + [b_i(x)d_i(x)]F'_i(x)\}. \end{aligned}$$

In summary, $\partial(\Phi_p)_i(x)^T \subseteq \{(D_a)_{ii}(x)e^i + \nabla F_i(x)(D_b)_{ii}(x)\}$ with $(D_a)_{ii}(x)$ and $(D_b)_{ii}(x)$ given as in the proposition. Thus, we complete the proof of part (c).

(d) For $i \in I_f$, then $(\Phi_p)_i(x) = -F_i(x)$. Obviously, $(D_a)_{ii}(x) = 0$, $(D_b)_{ii}(x) = -1$. \square

Proposition 3.2. For any given $x \in \mathbb{R}^n$, let $H \in \partial\Phi_p(x)$ with $H = D_a(x) + D_b(x)F'(x)$ where $D_a(x), D_b(x)$ are defined as in Lemma 3.1. Then we have

- (a) $[D_a(x)\Phi_p(x)]_i [D_b(x)\Phi_p(x)]_i \geq 0$ for all $i = 1, 2, \dots, n$.
- (b) $[D_a(x)\Phi_p(x)]_i = 0 \iff [D_b(x)\Phi_p(x)]_i = 0 \iff (\Phi_p)_i(x) = 0$ for each $i \notin I_f$.

Proof. (a) We discuss by the four cases as defined in Lemma 3.1 to complete the proof.

- (i) For $i \in I_l$, if $(x_i - l_i, F_i(x)) \neq (0, 0)$, then we have

$$\begin{aligned} &[D_a(x)\Phi_p(x)]_i [D_b(x)\Phi_p(x)]_i \\ &= \left(\frac{\text{sign}(x_i - l_i) \cdot |x_i - l_i|^{p-1}}{\|(x_i - l_i, F_i(x))\|_p^{p-1}} - 1 \right) \left(\frac{\text{sign}(F_i(x)) \cdot |F_i(x)|^{p-1}}{\|(x_i - l_i, F_i(x))\|_p^{p-1}} - 1 \right) [(\Phi_p)_i(x)]^2 \geq 0 \end{aligned}$$

where the inequality is due to the fact that

$$\left| \frac{\text{sign}(x_i - l_i) \cdot |x_i - l_i|^{p-1}}{\|(x_i - l_i, F_i(x))\|_p^{p-1}} \right| \leq 1 \quad \text{and} \quad \left| \frac{\text{sign}(F_i(x)) \cdot |F_i(x)|^{p-1}}{\|(x_i - l_i, F_i(x))\|_p^{p-1}} \right| \leq 1;$$

and if $(x_i - l_i, F_i(x)) = (0, 0)$, then we have

$$[D_a(x)\Phi_p(x)]_i [D_b(x)\Phi_p(x)]_i = (\xi - 1)(\zeta - 1) [(\Phi_p)_i(x)]^2 \geq 0$$

where the inequality is satisfied since $|\xi| \leq 1$ and $|\zeta| \leq 1$ when $|\xi|^{\frac{p}{p-1}} + |\zeta|^{\frac{p}{p-1}} \leq 1$.

- (ii) For $i \in I_u$, the verifications are the same as in case (i).

- (iii) For $i \in I_m$, from Lemma 3.1(c), we observe $a_i(x) \leq 0, b_i(x) \leq 0, c_i(x) \geq 0, d_i(x) \geq 0$. This immediately implies that $(D_a)_{ii}(x) \leq 0$ and $(D_b)_{ii}(x) \leq 0$. Therefore, $[D_a(x)\Phi_p(x)]_i [D_b(x)\Phi_p(x)]_i = (D_a)_{ii}(x)(D_b)_{ii}(x)[(\Phi_p)_i(x)]^2 \geq 0$.

- (iv) For $i \in I_f$, the result clearly holds since $[D_a(x)\Phi_p(x)]_i [D_b(x)\Phi_p(x)]_i = 0$.
- (b) Using part (a), [Lemma 3.1](#), and the same arguments as in [7, Lemma 3 (b)], the results can be verified. We omit the details due to the high similarity. \square

To know when all elements in the generalized Jacobian of Φ_p at a solution x^* of (1) are nonsingular, we introduce the concept of a strongly regular solution. Define

$$\begin{aligned}\alpha &:= \{i \in \{1, 2, \dots, n\} \mid l_i < x_i^* < u_i, F_i(x^*) = 0\}, \\ \beta &:= \{i \in \{1, 2, \dots, n\} \mid x_i^* \in \{l_i, u_i\}, F_i(x^*) = 0\}, \\ \gamma &:= \{i \in \{1, 2, \dots, n\} \mid x_i^* \in \{l_i, u_i\}, F_i(x^*) \neq 0\}.\end{aligned}\quad (22)$$

A solution x^* of (1) is said to be *strongly regular* if the submatrix $F'(x^*)_{\alpha\alpha}$ is nonsingular and the Schur complement

$$F'(x^*)_{\alpha\cup\beta, \alpha\cup\beta} / F'(x^*)_{\alpha\alpha} := F'(x^*)_{\beta\beta} - F'(x^*)_{\beta\alpha} F'(x^*)_{\alpha\alpha}^{-1} F'(x^*)_{\alpha\beta}$$

is a P -matrix. In addition, we also need the following technical lemma.

Lemma 3.2. Let $x^* \in \mathbb{R}^n$ be a solution of (1). Suppose that $H \in \partial\Phi_p(x^*)$ with $H = D_a(x^*) + D_b(x^*)F'(x^*)$ where $D_a(x^*)$, $D_b(x^*)$ are given as in [Lemma 3.1](#). Then,

- (a) $(D_a)_{ii}(x^*) = 0$ and $(D_b)_{ii}(x^*) = -1$ for all $i \in \alpha$.
 (b) $(D_a)_{ii}(x^*) \leq 0$, $(D_b)_{ii}(x^*) \leq 0$ and $(D_a)_{ii}(x^*) + (D_b)_{ii}(x^*) < 0$ for all $i \in \beta$.
 (c) $(D_a)_{ii}(x^*) = -1$ and $(D_b)_{ii}(x^*) = 0$ for all $i \in \gamma$.

Proof. The results are easily verified by [Lemma 3.1](#) and the definition in (22). \square

Proposition 3.3. Let $x^* \in \mathbb{R}^n$ be a strongly regular solution of (1). Then,

- (a) all elements $H \in \partial\Phi_p(x^*)$ are nonsingular;
 (b) there exists a constant $\kappa > 0$ such that $\|\Phi_p(x)\| \geq \kappa \|x - x^*\|$ for all $x \in \mathbb{R}^n$ in a sufficiently small neighborhood of x^* .

Proof. (a) Let $H \in \partial\Phi_p(x^*)$. We employ an idea used in [7, Theorem 1] to show that H is nonsingular, which is proving the homogeneous system $Hd = 0$ has only zero solution. From [Lemma 3.1](#), there exist diagonal matrices $D_a(x^*)$, $D_b(x^*) \in \mathbb{R}^{n \times n}$ such that

$$H = D_a(x^*) + D_b(x^*)F'(x^*). \quad (23)$$

We partition the vector d as $(d_\alpha, d_\beta, d_\gamma)$ and rearrange $D_a(x^*)$, $D_b(x^*)$ and $F'(x^*)$ as

$$\begin{aligned}D_a(x^*) &= \begin{pmatrix} (D_a)_{\alpha\alpha}(x^*) & 0 & 0 \\ 0 & (D_a)_{\beta\beta}(x^*) & 0 \\ 0 & 0 & (D_a)_{\gamma\gamma}(x^*) \end{pmatrix}, \\ D_b(x^*) &= \begin{pmatrix} (D_b)_{\alpha\alpha}(x^*) & 0 & 0 \\ 0 & (D_b)_{\beta\beta}(x^*) & 0 \\ 0 & 0 & (D_b)_{\gamma\gamma}(x^*) \end{pmatrix}, \\ F'(x^*) &= \begin{pmatrix} F'(x^*)_{\alpha\alpha} & F'(x^*)_{\alpha\beta} & F'(x^*)_{\alpha\gamma} \\ F'(x^*)_{\beta\alpha} & F'(x^*)_{\beta\beta} & F'(x^*)_{\beta\gamma} \\ F'(x^*)_{\gamma\alpha} & F'(x^*)_{\gamma\beta} & F'(x^*)_{\gamma\gamma} \end{pmatrix}.\end{aligned}$$

Applying [Lemma 3.2](#), the homogeneous system $Hd = 0$ can be recast as

$$F'(x^*)_{\alpha\alpha} d_\alpha + F'(x^*)_{\alpha\beta} d_\beta + F'(x^*)_{\alpha\gamma} d_\gamma = 0_\alpha, \quad (24)$$

$$(D_a)_{\beta\beta}(x^*) d_\beta + (D_b)_{\beta\beta}(x^*) [F'(x^*)_{\beta\alpha} d_\alpha + F'(x^*)_{\beta\beta} d_\beta + F'(x^*)_{\beta\gamma} d_\gamma] = 0_\beta, \quad (25)$$

$$-d_\gamma = 0_\gamma. \quad (26)$$

Since $d_\gamma = 0_\gamma$ and $F'(x^*)_{\alpha\alpha}$ is nonsingular by the assumption of x^* , from (24) we get

$$d_\alpha = -F'(x^*)_{\alpha\alpha}^{-1} F'(x^*)_{\alpha\beta} d_\beta. \quad (27)$$

Plugging $d_\gamma = 0_\gamma$ and d_α in (25) into (25) gives

$$[(D_a)_{\beta\beta}(x^*) + (D_b)_{\beta\beta}(x^*) (F'(x^*)_{\alpha\cup\beta, \alpha\cup\beta} / F'(x^*)_{\alpha\alpha})] d_\beta = 0_\beta. \quad (28)$$

Since the Schur complement $F'(x^*)_{\alpha\cup\beta, \alpha\cup\beta} / F'(x^*)_{\alpha\alpha}$ is a P -matrix by assumption and the diagonal matrices $(D_a)_{\beta\beta}(x^*)$, $(D_b)_{\beta\beta}(x^*)$ are negative semidefinite with a negative definite sum by [Lemma 3.2\(b\)](#), it follows from Proposition 2.7 of [27] that the coefficient matrix in (28) is nonsingular which implies $d_\beta = 0_\beta$. This together with (27) says $d_\alpha = 0_\alpha$. Hence, we can conclude that $d = 0$ which means H is nonsingular.

(b) Since all elements in $\partial\Phi_p(x^*)$ are nonsingular by part (a) and Φ_p is semismooth by [Proposition 3.1\(a\)](#), the desired result follows by [23, Proposition 3]. \square

The following proposition states that under some mild conditions a stationary point of the constrained minimization problem (17) is a solution of MCP (1).

Proposition 3.4. Let x^* be a stationary point of the reformulation (17) such that

- (a) the principal submatrix $\nabla F(x^*)_{I_f I_f}$ is nonsingular, and
- (b) the Schur complement $F'(x^*)/F'(x^*)_{I_f I_f} := F'(x^*)_{\bar{I}_f \bar{I}_f} - F'(x^*)_{\bar{I}_f I_f} F'(x^*)_{I_f I_f}^{-1} F'(x^*)_{I_f \bar{I}_f}$ is a P_0 -matrix, where $\bar{I}_f := \{1, 2, \dots, n\} \setminus I_f$.

Then x^* is a solution of (1).

Proof. We will complete the proof by the following two steps: (a) showing that every stationary point x^* of the constrained reformulation (17) is indeed a stationary point of (15); (b) further verifying that every stationary point of (15) is a solution of (1).

(a) Since x^* is a stationary point of the reformulation (17) of MCP (1), it satisfies

$$\begin{cases} x_i^* = l_i \Rightarrow [\nabla \Psi_p(x^*)]_i \geq 0, \\ x_i^* = u_i \Rightarrow [\nabla \Psi_p(x^*)]_i \leq 0, \\ x_i^* \in (l_i, u_i) \Rightarrow [\nabla \Psi_p(x^*)]_i = 0. \end{cases} \quad (29)$$

Suppose that x^* is not a stationary point of (15), i.e., $\nabla \Psi_p(x^*) \neq 0$. From Proposition 3.1(b) and Lemma 3.1, we know that $\nabla \Psi_p(x^*)$ can be expressed as

$$\nabla \Psi_p(x^*) = H^T \Phi_p(x^*) = D_a(x^*) \Phi_p(x^*) + \nabla F(x^*) D_b(x^*) \Phi_p(x^*) \quad (30)$$

for a matrix $H \in \partial \Phi_p(x^*)$ and certain diagonal matrices $D_a(x^*)$ and $D_b(x^*)$. The third implication of (29) says $[\nabla \Psi_p(x^*)]_f = 0_f$, which is equivalent to

$$[D_a(x^*) \Phi_p(x^*) + \nabla F(x^*) D_b(x^*) \Phi_p(x^*)]_{I_f} = 0_{I_f}. \quad (31)$$

For convenience, we write

$$\begin{aligned} D_a(x^*) &= \begin{pmatrix} (D_a)_{I_f I_f}(x^*) & 0 \\ 0 & (D_a)_{\bar{I}_f \bar{I}_f}(x^*) \end{pmatrix}, \\ D_b(x^*) &= \begin{pmatrix} (D_b)_{I_f I_f}(x^*) & 0 \\ 0 & (D_b)_{\bar{I}_f \bar{I}_f}(x^*) \end{pmatrix}, \\ \nabla F(x^*) &= \begin{pmatrix} \nabla F(x^*)_{I_f I_f} & \nabla F(x^*)_{I_f \bar{I}_f} \\ \nabla F(x^*)_{\bar{I}_f I_f} & \nabla F(x^*)_{\bar{I}_f \bar{I}_f} \end{pmatrix}. \end{aligned}$$

Noting that $(D_a)_{ii}(x^*) = 0$ and $(D_b)_{ii}(x^*) = -1$ for all $i \in I_f$ by Lemma 3.1(d), Eq. (31) is further equivalent to

$$-\nabla F(x^*)_{I_f I_f} \Phi_p(x^*)_{I_f} + \nabla F(x^*)_{I_f \bar{I}_f} (D_b)_{\bar{I}_f \bar{I}_f}(x^*) \Phi_p(x^*)_{\bar{I}_f} = 0_{I_f}. \quad (32)$$

Since $\nabla F(x^*)_{I_f I_f}$ is nonsingular by assumption, we can express (32) as

$$\Phi_p(x^*)_{I_f} = \nabla F(x^*)_{I_f I_f}^{-1} \nabla F(x^*)_{I_f \bar{I}_f} (D_b)_{\bar{I}_f \bar{I}_f}(x^*) \Phi_p(x^*)_{\bar{I}_f}. \quad (33)$$

Now plugging (33) into (30) for the expression $[\nabla \Psi_p(x^*)]_{\bar{I}_f}$ and using the fact that $(D_a)_{ii}(x^*) = 0$ and $(D_b)_{ii}(x^*) = -1$ for all $i \in I_f$, we obtain

$$\begin{aligned} [\nabla \Psi_p(x^*)]_{\bar{I}_f} &= (D_a)_{\bar{I}_f \bar{I}_f} \Phi_p(x^*)_{\bar{I}_f} - \nabla F(x^*)_{\bar{I}_f I_f} \Phi_p(x^*)_{I_f} + \nabla F(x^*)_{\bar{I}_f \bar{I}_f} (D_b)_{\bar{I}_f \bar{I}_f}(x^*) \Phi_p(x^*)_{\bar{I}_f} \\ &= (D_a)_{\bar{I}_f \bar{I}_f} \Phi_p(x^*)_{\bar{I}_f} + \left(\nabla F(x^*) / \nabla F(x^*)_{I_f I_f} \right) (D_b)_{\bar{I}_f \bar{I}_f}(x^*) \Phi_p(x^*)_{\bar{I}_f}. \end{aligned} \quad (34)$$

Since we assume $\nabla \Psi_p(x^*) \neq 0$, there exists an index $i \in \bar{I}_f$ such that either

$$x_i^* = l_i \quad \text{and} \quad [\nabla \Psi_p(x^*)]_i > 0 \quad (35)$$

or

$$x_i^* = u_i \quad \text{and} \quad [\nabla \Psi_p(x^*)]_i < 0. \quad (36)$$

However, from definition of Φ_p and Lemma 3.2, we observe that

(i) when $x_i^* = l_i$, $(D_b)_{ii}(x^*) \leq 0$ and, $(\Phi_p)_i(x^*) = \phi_p(x_i^* - l_i, F_i(x^*)) = |F_i(x^*)| - F_i(x^*) \geq 0$ if $i \in I_l$; $(\Phi_p)_i(x^*) = \phi_p(x_i^* - l_i, \phi_p(u_i - x_i^*, -F_i(x^*))) \geq 0$ if $i \in I_{lu}$. Therefore,

$$\left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \leq 0;$$

(ii) when $x_i^* = u_i$, $(D_b)_{ii}(x^*) \leq 0$ and, $(\Phi_p)_i(x^*) = -\phi_p(u_i - x_i^*, -F_i(x^*)) = -(|F_i(x^*)| + F_i(x^*)) \leq 0$ if $i \in I_u$; $(\Phi_p)_i(x^*) = \phi_p(x_i^* - l_i, \phi_p(u_i - x_i^*, -F_i(x^*))) \leq 0$ if $i \in I_{lu}$. Hence, we have

$$\left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \geq 0.$$

Premultiplying $[\nabla \Psi_p(x^*)]_i$ in (35) and (36) with $\left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i$ and using (34) and the last two equations, we obtain the following inequality

$$\begin{aligned} & \left[(D_a)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \\ & + \left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \left[\left(\nabla F(x^*) / \nabla F(x^*)_{\bar{l}_f \bar{l}_f} \right) (D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \leq 0 \end{aligned} \quad (37)$$

for all indices $i \in \bar{l}_f$ such that $[\nabla \Psi_p(x^*)]_i \neq 0$. In addition, from (34), it is clear that

$$\begin{aligned} & \left[(D_a)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i + \left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i \\ & \times \left[\left(\nabla F(x^*) / \nabla F(x^*)_{\bar{l}_f \bar{l}_f} \right) (D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_i = 0 \end{aligned}$$

for all indices $i \in \bar{l}_f$ such that $[\nabla \Psi_p(x^*)]_i = 0$.

Since $\nabla \Psi_p(x^*) \neq 0$ by assumption and $[\nabla \Psi_p(x^*)]_{\bar{l}_f} = 0$ by (29), we necessarily have $[\nabla \Psi_p(x^*)]_{\bar{l}_f} \neq 0$. This together with (34) and Proposition 3.2(b) implies that $(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \neq 0$. Thus, by the given condition that $\nabla F(x^*) / \nabla F(x^*)_{\bar{l}_f \bar{l}_f}$ is a P_0 -matrix, there exists an index $i_0 \in \bar{l}_f$ such that

$$\left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_{i_0} \neq 0, \quad (38)$$

and

$$\left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_{i_0} \left[\left(\nabla F(x^*) / \nabla F(x^*)_{\bar{l}_f \bar{l}_f} \right) (D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_{i_0} \geq 0. \quad (39)$$

Now using Proposition 3.2(a) and Eqs. (37)–(39) yields that

$$\left[(D_a)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_{i_0} \left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_{i_0} = 0,$$

which by Proposition 3.2(b) implies $\left[(D_b)_{\bar{l}_f \bar{l}_f}(x^*) \Phi_p(x^*)_{\bar{l}_f} \right]_{i_0} = 0$. This contradicts (38).

(b) Since we have proved $\nabla \Psi_p(x^*) = 0$, from Eq. (34) it follows that

$$\left[(D_a)_{\bar{l}_f \bar{l}_f}(x^*) + (\nabla F(x^*) / \nabla F(x^*)_{\bar{l}_f \bar{l}_f}) (D_b)_{\bar{l}_f \bar{l}_f}(x^*) \right] \Phi_p(x^*)_{\bar{l}_f} = 0_{\bar{l}_f}. \quad (40)$$

On the other hand, we observe that

- (i) the diagonal matrices $(D_a)_{\bar{l}_f \bar{l}_f}(x^*)$ and $(D_b)_{\bar{l}_f \bar{l}_f}(x^*)$ have nonpositive entries;
- (ii) a diagonal component of $(D_a)_{\bar{l}_f \bar{l}_f}(x^*)$ or $(D_b)_{\bar{l}_f \bar{l}_f}(x^*)$ can be zero only if the corresponding component of $\Phi_p(x^*)_{\bar{l}_f}$ is zero;
- (iii) the diagonal matrices $(D_a)_{\bar{l}_f \bar{l}_f}(x^*)$ and $(D_b)_{\bar{l}_f \bar{l}_f}(x^*)$ are always postmultiplied by $\Phi_p(x^*)_{\bar{l}_f}$ in the system (32) and (40);

we can assume without loss of generality that all diagonal entries of $D_a(x^*)$ and $D_b(x^*)$ are negative. Then Lemma 2.3 and our assumption for this proposition show that the coefficient matrix in (40) is nonsingular which says $\Phi_p(x^*)_{\bar{l}_f} = 0_{\bar{l}_f}$. This further implies $\Phi_p(x^*)_f = 0_f$ by (33). Hence, $\Phi_p(x^*) = 0$, i.e., x^* is a solution of MCP (1). \square

3.2. Algorithm and convergence results

Now we describe the strictly feasible Newton-type method used in [8] to solve the constrained minimization problem (17). The detailed iterative scheme is as follows.

Algorithm 3.1 (Strictly Feasible Newton-type Method).

(S.0) Choose $x^0 \in (l, u)$, $\delta > 0$, $c > 0$, $\tau \in (0, 1)$, $\omega \in (0, 1)$, $\rho > 0$, $p_1 > 1$, $p_2 > 0$, $\beta \in (0, 1)$, $\sigma \in (0, 1)$, $\gamma > 0$, and set $k := 0$.

(S.1) If x^k is a stationary point of (15), then stop.

(S.2) Define $\delta_k := \min \left\{ \delta, c\sqrt{\|\Phi_p(x^k)\|} \right\}$ and

$$\mathcal{A}_k := \{i \mid x_i^k - l_i \leq \delta_k \text{ or } u_i - x_i^k \leq \delta_k\}, \quad \mathcal{I}_k := \{1, 2, \dots, n\} \setminus \mathcal{A}_k.$$

(S.3) Select an element $H^k \in \partial_C \Phi_p(x^k)$ and compute a direction $d^k \in \mathbb{R}^n$ in the following way. For $i \in \mathcal{A}_k$, set

$$d_i^k := \begin{cases} l_i - x_i^k & \text{if } x_i^k - l_i \leq \delta_k, \\ u_i - x_i^k & \text{if } u_i - x_i^k \leq \delta_k, \end{cases}$$

then solve the linear system

$$H_{\mathcal{I}_k \mathcal{I}_k}^k d_{\mathcal{I}_k} = -\Phi_p(x^k)_{\mathcal{I}_k} - H_{\mathcal{I}_k \mathcal{A}_k}^k d_{\mathcal{A}_k}^k \quad (41)$$

in order to obtain components d_i^k for $i \in \mathcal{I}_k$. If the system (41) is not solvable, then set $\tau_k := \tau$ and go to (S.6), otherwise go to (S.4).

(S.4) Set $\tau_k := \max\{\tau, 1 - \|\Phi_p(x^k)\|\}$. If $x^k + \tau_k d^k \in (l, u)$ and

$$\|\Phi_p(x^k + \tau_k d^k)\| \leq \omega \|\Phi_p(x^k)\|, \quad (42)$$

then set $x^{k+1} := x^k + \tau_k d^k$ and go to (S.7), else go to (S.5).

(S.5) Set $\bar{x}_N^k := [x^k + d^k]_+$ and $s_N^k := \bar{x}_N^k - x^k$. If

$$\nabla \Psi_p(x^k)^T s_N^k \leq -\rho \|s_N^k\|^{p_1} \quad \text{and} \quad \nabla \Psi_p(x^k)^T s_N^k \leq -\rho \|\Phi_p(x^k)\|^{p_2},$$

then compute $t_k := \max\{\tau_k \beta^l \mid l = 0, 1, 2, \dots\}$ such that

$$\Psi_p(x^k + t_k s_N^k) \leq \Psi_p(x^k) + \sigma t_k \nabla \Psi_p(x^k)^T s_N^k,$$

set $x^{k+1} := x^k + t_k s_N^k$ and go to (S.7), otherwise go to (S.6).

(S.6) Set $\bar{x}_G^k := [x^k - \gamma \nabla \Psi_p(x^k)]_+$ and $s_G^k := \bar{x}_G^k - x^k$. Compute $t_k := \max\{\tau_k \beta^l \mid l = 0, 1, 2, \dots\}$ such that

$$\Psi_p(x^k + t_k s_G^k) \leq \Psi_p(x^k) + \sigma t_k \nabla \Psi_p(x^k)^T s_G^k,$$

set $x^{k+1} := x^k + t_k s_G^k$ and go to (S.7).

(S.7) Set $k \leftarrow k + 1$, and go to (S.1).

With the aid of the properties obtained in Section 3.1, all the convergence results in Section 4 of [8] can be verified for Algorithm 3.1. We only summarize them and omit the detailed arguments since their proofs are similar.

Theorem 3.1. (a) Algorithm 3.1 is well defined, and particularly we have $\{x^k\} \subseteq (l, u)$.

(b) Every accumulation point of $\{x^k\}$ generated by Algorithm 3.1 is a stationary point of (17).

(c) If the conditions of Proposition 3.4 are satisfied at x^* , then the accumulation point x^* is a solution of MCP (1).

Theorem 3.2. Assume that the accumulation point x^* of sequence $\{x^k\}$ generated by Algorithm 3.1 is a strongly regular solution for MCP (1). Then,

(a) the entire sequence $\{x^k\}$ converges Q -superlinearly to x^* ;

(b) if F' is locally Lipschitzian around x^* , the rate of convergence is Q -quadratic.

3.3. Numerical experiments

We implemented Algorithm 3.1 in MATLAB 7.0 for solving the MCPLIB test problems [11]. The actual implementation differs slightly from the description of Algorithm 3.1. Similar to [8], for Step (S.5), we adopted the nonmonotone line search in [28] instead of the monotone line search, i.e., we computed t_k such that

$$\Psi_p(x^k + t_k s_N^k) \leq \mathcal{W}_k + \sigma t_k \nabla \Psi_p(x^k)^T s_N^k$$

where

$$\mathcal{W}_k := \max \{ \Psi_p(x^j) \mid j = k + 1 - m_k, \dots, k \}$$

denotes the maximum function value of Ψ_p over the last m_k iterations. During our tests, we set $m_k = 1$ for $k = 0, 1, \dots, 5$ and $m_{k+1} = \min\{m_k + 1, m\}$ with $m = 5$ for all remaining iterations. In addition, we also adopted the so-called *watchdog strategy* to enhance Step (S.5) of Algorithm 3.1. If after 10 steps the best function value of Ψ_p found so far has not been reduced sufficiently, we return to that point using a monotone line search.

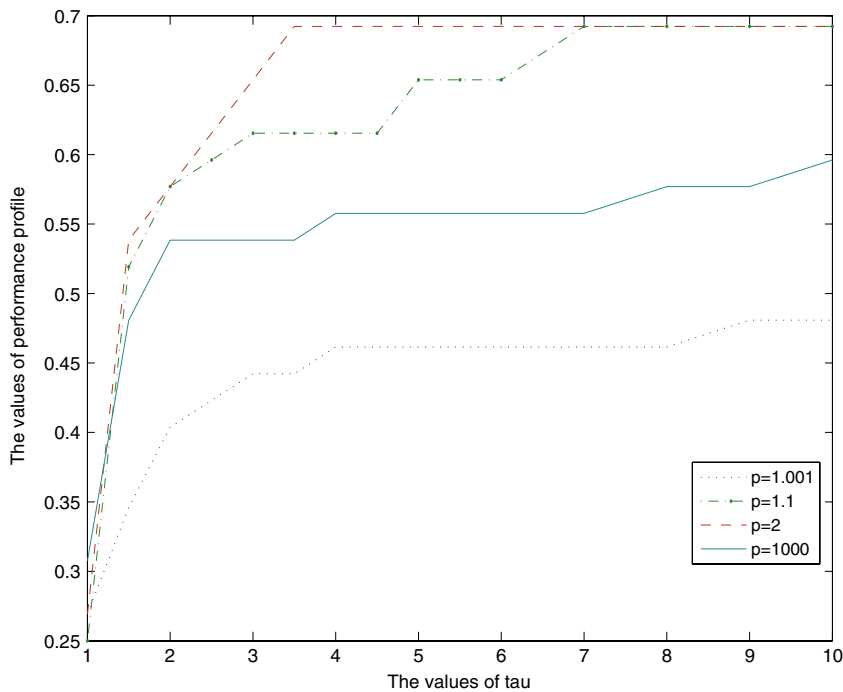


Fig. 1. Performance profile of iterations for Algorithm 3.1 with four p on the MCPLIB collection.

All experiments were done with a PC of Intel Pentium Dual CPU E2200 and 2048MB memory. The parameters of Algorithm 3.1 were chosen as follows:

$$\begin{aligned} \beta &= 0.5, & \sigma &= 10^{-4}, & \delta &= \sigma = 10^{-4}, & \tau &= 0.95, & \omega &= 0.995, \\ \rho &= 10^{-12}, & p_1 &= 2.1, & p_2 &= c = \gamma = 1. \end{aligned} \quad (43)$$

We started Algorithm 3.1 with a strictly feasible point $x^0 = \max\{l + e, \min\{\hat{x}, u\} - e\}$, where $e \in \mathbb{R}^n$ is a vector with all components being 1 and \hat{x} is the standard starting point provided by the MCPLIB collection. We terminated the iteration whenever

$$\|\Phi_p(x)\| \leq 10^{-11} \quad \text{and} \quad \|\nabla \Psi_p(x)\| \leq 10^{-3}, \quad \text{or} \quad \|\nabla \Psi_p(x)\| \leq 5.0 \times 10^{-7}, \quad \text{or} \quad k > 500.$$

The procedure for calculating an element $H_k \in \partial_C \Phi_p(x^k)$ is similar to the one given in [29]. To present an objective evaluation and comparison of the performance of Algorithm 3.1 with different p , we adopt the performance profile introduced in [20] as a means.

Specifically, we regard Algorithm 3.1 corresponding to a p as a **solver**, and assume that there are n_s solvers and n_j test problems from the MCPLIB collection \mathcal{J} . We are interested in using the number of iterations and function evaluations as two performance measures for Algorithm 3.1 with different p . For each problem j and solver s , let

$$\begin{aligned} k_{j,s} &:= \text{the iterations required to solve problem } j \text{ by solvers}, \\ f_{j,s} &:= \text{function evaluations required to solve problem } j \text{ by solvers}. \end{aligned}$$

We compare the performance on problem j by solver s with the best performance by any one of the n_s solvers on this problem; that is, we adopt the **performance ratio**

$$r_{j,s} = \frac{k_{j,s}}{\min\{k_{j,s} : s \in \mathcal{S}\}} \quad \text{or} \quad \frac{f_{j,s}}{\min\{f_{j,s} : s \in \mathcal{S}\}},$$

where \mathcal{S} is the set of four solvers. An overall assessment of each solver is obtained from

$$\rho_s(\tau) = \frac{1}{n_j} \text{size} \{j \in \mathcal{J} : r_{j,s} \leq \tau\}, \quad (44)$$

which is called the **performance profile** of the number of iterations (or function evaluations) for solver s . Note that $\rho_s(\tau)$ approximates the probability for solver $s \in \mathcal{S}$ that a performance ratio $r_{j,s}$ is within a factor $\tau \in \mathbb{R}$ of the best possible ratio.

Fig. 1 shows the performance profile of iterations in the range of $[0, 10]$ for four solvers on 52 test problems. The four solvers correspond to Algorithm 3.1 with $p = 1.001$, $p = 1.1$, $p = 2$, and $p = 1000$, respectively. From this figure, we

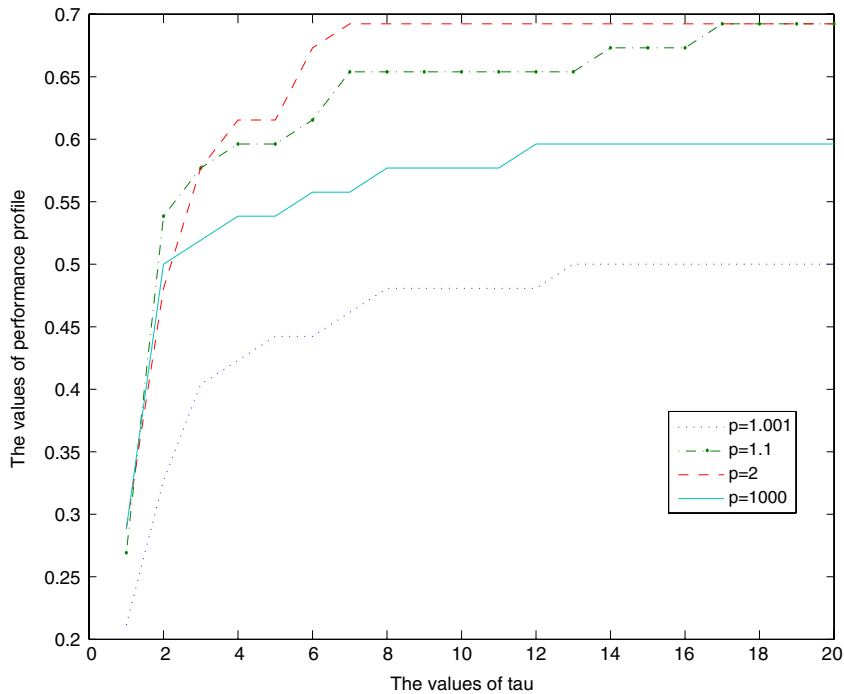


Fig. 2. Performance profile of function evaluations for Algorithm 3.1 with four p on the MCPLIB collection.

see that Algorithm 3.1 with $p = 1000$ has the most wins (has the highest probability of being the optimal solver) and that the probability that it is the winner on a given MCP is about 0.31. If we choose being within a factor of 2 or 7 of the best solver as the scope of our interest, then either $p = 1.1$ or $p = 2$ would suffice, and the performance profile shows that the probability that Algorithm 3.1 with the two p can solve a given MCP within a factor 2 of the best solver is about 58%, and the probability that they can solve a given MCP within a factor 7 of the best solver is enhanced to 70%. Although $p = 1.001$ has a competitive number of wins with $p = 1.1$ and $p = 2$, it is not a good choice since the probability that it can solve a given MCP within any factor of the best solver is the lowest.

Fig. 2 shows the performance profile of function evaluations in the range of $[0, 20]$ for the above four solvers on the same 52 test problems. From this figure, we see that Algorithm 3.1 with $p = 2$ and $p = 1000$ has the competitive wins and that the probability that it is the winner on a given MCP is about 0.28. If we choose being within a factor of greater than 2 of the best solver as the scope of our interest, then either $p = 1.1$ or $p = 2$ would suffice, and the performance profile shows that the probability that Algorithm 3.1 with the two p can solve a given MCP in such range of the best solver is over 50%, and it may increase to 70% within a factor 17 of the best solver. Although $p = 1000$ has a competitive number of wins with $p = 2$, the probability that it can solve a given MCP within any positive factor of the best solver is lower than $p = 2$. In addition, it is clear that Algorithm 3.1 with $p = 1.001$ is the worst choice among the four solvers.

To sum up, Algorithm 3.1 with $p = 1.1$ and $p = 2$ have the best performance whether by iterations or function evaluations within any positive factor of the best solver, whereas Algorithm 3.1 with $p = 1.001$ has the worst performance whether by iterations or function evaluations. Although Algorithm 3.1 with $p = 1000$ tends to have the highest probability of being the optimal solver for a given MCP problem, but it has a lower probability than $p = 1.1$ and $p = 2$ within any positive factor of the best solver.

4. Semismooth Levenberg–Marquardt method

In this section, we study the semismooth Levenberg–Marquardt method based on the generalized FB function, and extend the convergence results in [9] which used the FB function to this case. To this end, we define an operator $\Phi_p : \mathbb{R}^n \rightarrow \mathbb{R}^{2n}$ componentwise by

$$(\Phi_p)_i(x) = \begin{cases} \lambda \phi_p(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ -\lambda \phi_p(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ \lambda \phi_p(x_i - l_i, \phi_p(u_i - x_i, -F_i(x))) & \text{if } i \in I_{lu}, \\ -\lambda F_i(x) & \text{if } i \in I_f, \end{cases} \quad (45)$$

$$(\bar{\Phi}_p)_{n+i}(x) = \begin{cases} (1-\lambda)\phi_+(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ (1-\lambda)\phi_+(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ (1-\lambda)\left(\phi_+(x_i - l_i, F_i(x)) + \phi_+(u_i - x_i, -F_i(x))\right) & \text{if } i \in I_{lu}, \\ -(1-\lambda)F_i(x) & \text{if } i \in I_f, \end{cases} \quad (46)$$

where $\phi_+(a, b) := a_+ b_+$. It is not hard to verify that MCP (1) is also equivalent to the overdetermined system of equations

$$\bar{\Phi}_p(x) = 0, \quad (47)$$

and the squared norm of the operator $\bar{\Phi}_p$ then induces a family of merit functions in the sense that the solution of MCP (1) is equivalent to finding a minimizer of the following unconstrained minimization problem whose objective value equals 0:

$$\min_{x \in \mathbb{R}^n} \bar{\Psi}_p(x) := \frac{1}{2} \|\bar{\Phi}_p(x)\|^2. \quad (48)$$

4.1. Properties of $\bar{\Phi}_p$ and $\bar{\Psi}_p$

In what follows, we present some favorable properties of $\bar{\Phi}_p$ and $\bar{\Psi}_p$ which are useful in the subsequent analysis although their proofs are routine. Since the proof of Proposition 4.1 is direct by [17, Propositions 3.1 and 3.3] and [17, Lemma 2.3], we omit it.

Proposition 4.1. Let $\bar{\Phi}_p$ be defined by (45)–(46) and $\bar{\Psi}_p$ be given by (48). Then,

- (a) The mapping $\bar{\Phi}_p$ is semismooth, and moreover, it is strongly semismooth if F' is Lipschitz continuous.
- (b) The function $\bar{\Psi}_p$ is continuously differentiable everywhere.

In Section 3.1, an overestimation of the generalized Jacobian $\partial \Phi_p(x^*)$, which is actually the C-subdifferential $\partial_C \Phi_p(x^*)$, is given and a so-called strongly regular condition is used to describe when all the elements of $\partial \Phi_p(x^*)$ are nonsingular. Here, we will present the expression of the C-subdifferential $\partial_C \bar{\Phi}_p(x^*)$, and under the same condition show that all the elements of $\partial_C \bar{\Phi}_p(x^*)$ have full rank which will be employed as an assumption to guarantee the local convergence of the algorithm studied in Section 4.2 later.

Proposition 4.2. Let $\bar{\Phi}_p : \mathbb{R}^n \rightarrow \mathbb{R}^{2n}$ be defined by (45)–(46). The following results hold.

- (a) Any matrix $H \in \partial_C \bar{\Phi}_p(x)$ has the representation $H = \begin{pmatrix} \lambda H_1 \\ (1-\lambda)H_2 \end{pmatrix}$, with $H_1 \subseteq D_a(x) + D_b(x)F'(x)$ and $H_2 \subseteq \tilde{D}_a(x) + \tilde{D}_b(x)F'(x)$ where $D_a(x), D_b(x) \in \mathbb{R}^{n \times n}$ are diagonal matrices whose diagonal elements are given by Lemma 3.1, and $\tilde{D}_a(x), \tilde{D}_b(x)$ are $n \times n$ diagonal matrices whose diagonal entries are given below.
 - (i) If $i \in I_l$, then $(\tilde{D}_a)_{ii}(x) = F_i(x)_+ \partial(x_i - l_i)_+$ and $(\tilde{D}_b)_{ii}(x) = (x_i - l_i)_+ \partial F_i(x)_+$.
 - (ii) If $i \in I_u$, then

$$(\tilde{D}_a)_{ii}(x) = (-F_i(x))_+ \partial(u_i - x_i)_+, \quad (\tilde{D}_b)_{ii}(x) = (u_i - x_i)_+ \partial(-F_i(x))_+.$$
 - (iii) If $i \in I_{lu}$, then

$$(\tilde{D}_a)_{ii}(x) = F_i(x)_+ \partial(x_i - l_i)_+ + (-F_i(x))_+ \partial(u_i - x_i)_+,$$

$$(\tilde{D}_b)_{ii}(x) = (x_i - l_i)_+ \partial F_i(x)_+ + (u_i - x_i)_+ \partial(-F_i(x))_+.$$
 - (iv) If $i \in I_f$, then $(\tilde{D}_a)_{ii}(x) = 0, (\tilde{D}_b)_{ii}(x) = -1$.
- (b) Suppose that $x^* \in \mathbb{R}^n$ is a strongly regular solution of (1). Then, all elements from $\partial_C \bar{\Phi}_p(x^*)$ have full rank.

Proof. (a) From Eq. (8), $(\partial_C \bar{\Phi}_p(x))^T = \partial(\bar{\Phi}_p)_1(x) \times \cdots \times \partial(\bar{\Phi}_p)_{2n}(x)$, where $\partial(\bar{\Phi}_p)_i(x)$ denotes the generalized gradient of the i th component function of $\bar{\Phi}_p$. From Lemma 3.1, it follows that $\partial(\bar{\Phi}_p)_i(x) \subseteq \lambda((D_a)_{ii}(x)(e^i)^T + (D_b)_{ii}(x)F'_i(x))$ for $i = 1, \dots, n$, where $D_a(x), D_b(x) \in \mathbb{R}^{n \times n}$ are diagonal matrices whose diagonal elements are given as in Lemma 3.1. Since the generalized gradient of the function $\phi_+ : \mathbb{R}^2 \rightarrow \mathbb{R}$ at a point $(a, b) \in \mathbb{R}^2$ is equal to $\partial \phi_+(a, b) = \{(b_+ \partial a_+, a_+ \partial b_+)\}$ where

$$\partial z_+ = \begin{cases} 1 & \text{if } z > 0, \\ [0, 1] & \text{if } z = 0, \\ 0 & \text{if } z < 0, \end{cases}$$

we have $\partial(\bar{\Phi}_p)_{n+i}(x) \subseteq (1-\lambda)((\tilde{D}_a)_{ii}(x)(e^i)^T + (\tilde{D}_b)_{ii}(x)F'_i(x))$ for each i . Specifically, we observe

- (i) if $i \in I_l$, $(\bar{\Phi}_p)_{n+i}(x) = (1-\lambda)\phi_+(x_i - l_i, F_i(x))$;
- (ii) if $i \in I_u$, $(\bar{\Phi}_p)_{n+i}(x) = (1-\lambda)\phi_+(u_i - x_i, -F_i(x))$;

- (iii) if $i \in I_{lu}$, $(\bar{\Phi}_p)_{n+i}(x) = (1 - \lambda)(\phi_+(x_i - l_i, F_i(x)) + \phi_+(u_i - x_i, -F_i(x)))$;
 (iv) if $i \in I_f$, $(\bar{\Phi}_p)_{n+i}(x) = -(1 - \lambda)F_i(x)$ and clearly $(\bar{D}_a)_{ii}(x) = 0$, $(\bar{D}_b)_{ii}(x) = -1$.

From all the above observations and the generalized gradient of ϕ_+ , the desired result follows.

- (b) Let $H \in \partial_C(\bar{\Phi}_p)(x^*)$. By part (a), $H = \begin{pmatrix} \lambda H_1 \\ (1 - \lambda)H_2 \end{pmatrix}$, where H_1 is an element from $\partial_C(\Phi_p)(x^*)$. Since x^* is strongly regular, using the similar arguments as in Proposition 3.3(a) yields that $H_1 \in \partial_C(\Phi_p)(x^*)$ is nonsingular, which implies $\text{rank}(H) = n$. \square

4.2. Algorithm and convergence results

Algorithm 4.1 (Semismooth Levenberg–Marquardt Method).

(S.0) Choose $x^0 \in \mathbb{R}^n$, $\lambda \in (0, 1)$, $\beta \in (0, 1)$, $\sigma \in (0, 1/2)$ and $\varepsilon \geq 0$. Set $k := 0$.

(S.1) If $\|\nabla \bar{\Psi}_p(x^k)\| \leq \varepsilon$, then stop.

(S.2) Choose $H_k \in \partial_C \bar{\Phi}_p(x^k)$ and $\nu_k > 0$. Find a solution $d^k \in \mathbb{R}^n$ of linear system

$$(H_k^T H_k + \nu_k I) d = -\nabla \bar{\Psi}_p(x^k), \quad (49)$$

where $\nu_k > 0$ is the Levenberg–Marquardt parameter.

(S.3) Compute $t_k = \max\{\beta^l \mid l = 0, 1, 2, \dots\}$ such that

$$\bar{\Psi}_p(x^k + t_k d^k) \leq \bar{\Psi}_p(x^k) + \sigma t_k \nabla \bar{\Psi}_p(x^k)^T d^k, \quad (50)$$

and let $x^{k+1} := x^k + t_k d^k$.

(S.4) Set $k := k + 1$, and go to (S.1).

Notice that the above method is different from the classical Levenberg–Marquardt method for nonlinear least-square problems in which $\bar{\Phi}_p$ is not continuously differentiable. If $\nu_k \equiv 0$, the solution of (49) is exactly the solution of the linear least-square problem

$$\min_{d \in \mathbb{R}^n} \frac{1}{2} \|H_k d + \bar{\Phi}_p(x^k)\|^2, \quad (51)$$

since $\bar{\Psi}_p(x)$ is continuously differentiable and $\nabla \bar{\Psi}_p(x^k) = H_k^T \bar{\Phi}_p(x^k)$. In this paper, we choose the Levenberg–Marquardt parameter ν_k by

$$\nu_k := \min\{\rho_1, \rho_2 \|\bar{\Phi}_p(x^k)\|^\varrho\} \quad (52)$$

for some constants $\rho_1, \rho_2 > 0$, where $\varrho \in [1, 2]$. Such choice is not only consistent with the requirements for local superlinear (quadratic) convergence stated in Theorem 4.2, but also is adopted in our numerical experiments later.

In what follows, we will study the convergence properties of the algorithm. For this purpose, we assume that $\varepsilon = 0$. The first one is a global convergence result.

Theorem 4.1. Let $\{x^k\}$ be the sequence generated by Algorithm 4.1. Then,

- (a) $\{x^k\}$ is well defined, and every accumulation point is a stationary point of (48).
 (b) If x^* is a stationary point of (48) such that conditions (a) and (b) of Proposition 3.4 hold, then x^* is a solution of MCP (1).

Proof. (a) From the steps of Algorithm 4.1, $\{x^k\}$ is well defined since $\nu_k > 0$, and d^k determined by (49) is always a descent direction of $\bar{\Psi}_p$ at x^k . Let x^* be any accumulation point of $\{x^k\}$ and $\{x^k\}_K$ be a subsequence converging to x^* . Suppose that $\nabla \bar{\Psi}_p(x^*) \neq 0$. Since $\{\bar{\Psi}_p(x^k)\}$ is monotonically decreasing and bounded below, and $\{\bar{\Psi}_p(x^k)\}_K$ converges to $\bar{\Psi}_p(x^*)$, the entire sequence $\{\bar{\Psi}_p(x^k)\}$ converges to $\bar{\Psi}_p(x^*) > 0$. Since

$$\bar{\Psi}_p(x^{k+1}) - \bar{\Psi}_p(x^k) \leq \sigma t_k \nabla \bar{\Psi}_p(x^k)^T d^k \leq 0$$

for all sufficiently large k , using $\bar{\Psi}_p(x^{k+1}) - \bar{\Psi}_p(x^k) \rightarrow 0$ yields that

$$\{t_k \nabla \bar{\Psi}_p(x^k)^T d^k\}_K \rightarrow 0. \quad (53)$$

We next prove $\{\nabla \bar{\Psi}_p(x^k)^T d^k\}_K$ has a nonzero limit as $k \rightarrow +\infty$. By the definition of d^k ,

$$\nabla \bar{\Psi}_p(x^k)^T d^k = -\nabla \bar{\Psi}_p(x^k)^T (H_k^T H_k + \nu_k I)^{-1} \nabla \bar{\Psi}_p(x^k) \quad \forall k. \quad (54)$$

Since the C-subdifferential $\partial_C \bar{\Phi}_p(x)$ is a nonempty compact set for any $x \in \mathbb{R}^n$, $\{H_k\}_K$ is bounded. Without loss of generality, we assume that $\{H_k\}_K \rightarrow H_*$. Considering that the set-valued mapping $x \mapsto \partial_C \bar{\Phi}_p(x)$ is closed and $\{x^k\}_K \rightarrow x^*$, we have $H_* \in \partial_C \bar{\Phi}_p(x^*)$. In addition, by the continuity of $\bar{\Phi}_p$ and $\bar{\Phi}_p(x^*) \neq 0$, we have $\nu_k \rightarrow \nu_*$ with $\nu_* = \min\{\rho_1, \rho_2 \|\bar{\Phi}_p(x^*)\|^\varrho\} > 0$.

Thus, $\{H_k^T H_k + \nu_k I\}_{k \in K} \rightarrow H_*^T H_* + \nu_* I \succ 0$. This, together with (54) and the continuity of $\nabla \bar{\Psi}_p$, implies that $\{\nabla \bar{\Psi}_p(x^k)^T d^k\}_K$ has a nonzero limit as $k \rightarrow +\infty$. From (53), we then obtain $\{t_k\}_K \rightarrow 0$.

Now, for all sufficiently large k , let $l_k \in \{0, 1, \dots\}$ be the unique exponent such that $t_k = \beta^{l_k}$. Since $\{t_k\}_K \rightarrow 0$, we have $\{l_k\}_{k \in K} \rightarrow \infty$. From the Armijo line search in (S.3),

$$\frac{\bar{\Psi}_p(x^k + \beta^{l_k-1} d^k) - \bar{\Psi}_p(x^k)}{\beta^{l_k-1}} > \sigma \nabla \bar{\Psi}_p(x^k)^T d^k \quad (55)$$

for all $k \in K$ sufficiently large. Taking the limit $k \rightarrow \infty$ with $k \in K$ and using $\{l_k\}_K \rightarrow \infty$ and $\{x^k\}_K \rightarrow x^*$, we have $\nabla \bar{\Psi}_p(x^*)^T d^* \geq \sigma \nabla \bar{\Psi}_p(x^*)^T d^*$. This means $\nabla \bar{\Psi}_p(x^*)^T d^* \geq 0$. On the other hand, we learn from (49) that $\{d^k\}_K \rightarrow d^*$ with d^* being the solution of

$$(H_*^T H_* + \nu_* I) d = -\nabla \bar{\Psi}_p(x^*). \quad (56)$$

This means $\nabla \bar{\Psi}_p(x^*)^T d^* < 0$ since $(H_*^T H_* + \nu_* I) \succ 0$. Thus, we have a contradiction.

Using the same arguments as Proposition 3.4 yields the desired result. \square

By previous propositions and mimicking the arguments as in [9], we can obtain the following results.

Theorem 4.2. Let $\{x^k\}$ be the sequence generated by Algorithm 4.1. Assume that the accumulation point x^* of $\{x^k\}$ is a strongly regular solution of MCP (1). Then,

- (a) the entire sequence $\{x^k\}$ converges to x^* ;
- (b) the full stepsize $t_k = 1$ is always acceptable for k sufficiently large and the rate of convergence is Q -superlinear;
- (c) if, in addition, F' is locally Lipschitzian around x^* and $\nu_k = O(\|\nabla \bar{\Psi}_p(x^k)\|)$, the rate of convergence is Q -quadratic.

4.3. Numerical experiments

In this subsection, we report numerical results with Algorithm 4.1 solving the MCPLIB collection [11]. All experiments were done with a PC of Intel Pentium Dual CPU E2200 and 2048MB memory, and the computer codes were written in Matlab 7.0. The implementation of the algorithm is along the lines of Algorithm 4.1 except that the monotone line search in (S.3) was replaced by the nonmonotone line search proposed in [28]. In other words, we computed t_k such that $\bar{\Psi}_p(\zeta^k + t_k d^k) \leq \mathcal{W}_k + \sigma t_k \nabla \bar{\Psi}_p(\zeta^k)^T d^k$, where

$$\mathcal{W}_k := \max \{ \bar{\Psi}_p(\zeta^j) \mid j = k+1 - m_k, \dots, k \}$$

denotes the maximum function value of $\bar{\Psi}_p$ over the last m_k iterations. During our tests, we employed the same strategy as in [9] to adjust m_k . That is, $m_k = 1$ for $k = 0, 1, \dots, 5$ and $m_{k+1} = \min\{m_k + 1, m\}$ with $m = 10$ for all remaining iterations. In addition, we also adopted the so-called *watchdog strategy* to enhance Algorithm 4.1. If after 20 steps the best function value of $\bar{\Psi}_p$ found so far has not been reduced sufficiently, we return to that point using a monotone line search.

We started Algorithm 4.1 with the standard starting point provided by the MCPLIB collection, and terminated the iteration if one of the following conditions are satisfied

$$\|\bar{\Phi}_p(x)\| \leq 10^{-11} \quad \text{and} \quad \|\nabla \bar{\Psi}_p(x)\| \leq 10^{-4}, \quad \text{or} \quad k > 300.$$

The Levenberg–Marquardt parameter ν_k is chosen as follows: For smaller problems with $n < 100$, we first estimate the condition number of the matrix $H_k^T H_k$. If this estimated condition number is larger than 10^{25} , we set $\nu_k := 0.1/(k+1)$, and otherwise we set ν_k as in (52) with $\rho_1 = 1$, $\rho_2 = 5.0 \times 10^{-7}/n$, and $\varrho = 1$. For those problems with $n > 100$, we let $\nu_k = 0$. The other parameters in Algorithm 4.1 were chosen as follows:

$$\lambda = 0.9, \quad \beta = 0.55, \quad \sigma = 10^{-4}.$$

The procedure for calculating an element $H_k \in \partial_C \bar{\Psi}_p(x^k)$ is similar to the one given in [29].

Fig. 3 shows the performance profile of iteration times (defined as in (44)) in the range of $[0, 10]$ for four solvers on 55 test problems. The four solvers correspond to Algorithm 4.1 with $p = 1.001$, $p = 1.1$, $p = 2$, and $p = 1000$, respectively. From Fig. 3, it is clear that Algorithm 4.1 with $p = 1.001$ has the most wins (has the highest probability of being the optimal solver) and that the probability that it is the winner on a given MCP is about 0.58, and furthermore, it has the highest probability within any positive factor of the best solver. The performance profile shows that Algorithm 4.1 with $p = 1.1$ has a competitive performance with $p = 1.001$ if we choose being within a factor of greater than 4 of the best solver as the scope of our interest, though it has the lowest number of wins. Algorithm 4.1 with $p = 1000$ has the lowest probability within a factor of greater than 2 of the best solver, although it has higher number of wins than either $p = 1.1$ or $p = 2$. Comparing with Fig. 1, it is easy to see that for these p , the lowest probability of Algorithm 4.1 is higher than the highest probability of Algorithm 3.1 if we choose being within a factor of greater than 2 of the best solver as the scope of our interest.

Fig. 4 shows the performance profile of function evaluations in the range of $[0, 20]$ for the above four solvers on 55 test problems. From this figure, it is clear that Algorithm 4.1 with $p = 1.001$ also has the most wins in terms of function

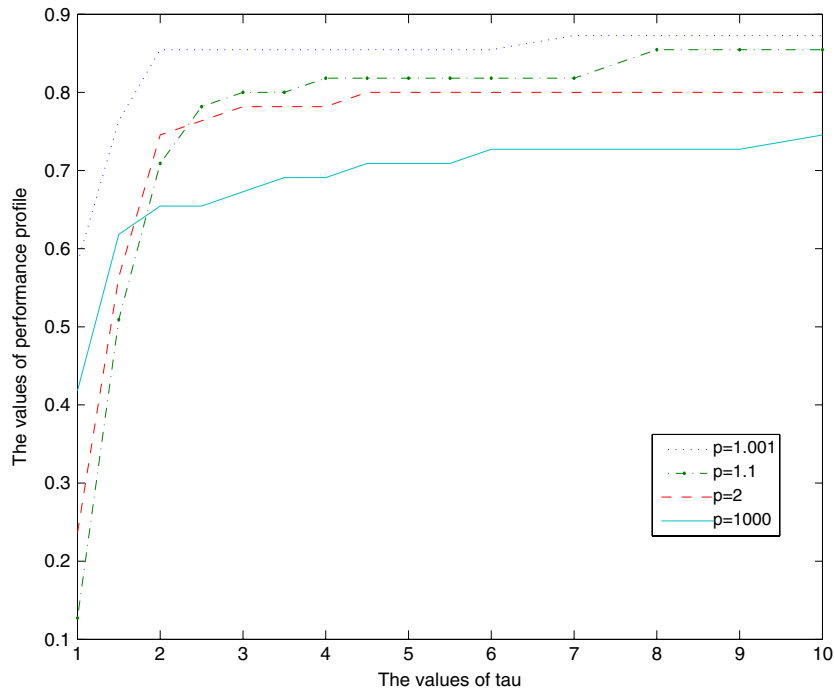


Fig. 3. Performance profile of iterations for Algorithm 4.1 with four p on the MCPLIB collection.

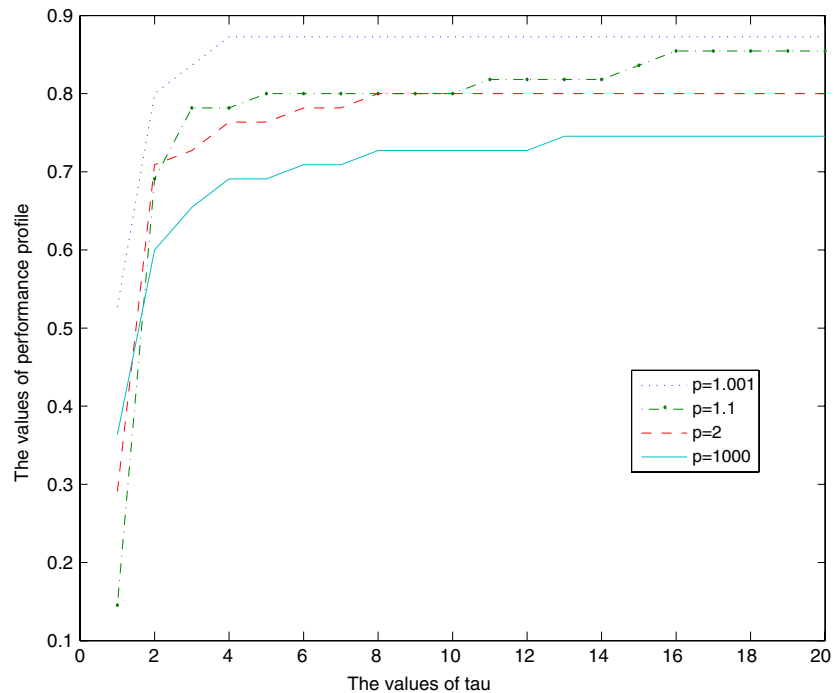


Fig. 4. Performance profile of function evaluations for Algorithm 4.1 with four p on the MCPLIB collection.

evaluations and that the probability that it is the winner on a given MCP is about 0.53, and furthermore, it has the highest probability within any positive factor of the best solver. The performance profile shows that Algorithm 4.1 with $p = 1.1$ has a comparable performance with $p = 1.001$ if we choose being within a factor of 16 of the best solver as the scope of our interest, though it has the lowest number of wins. Algorithm 4.1 with $p = 1000$ has the lowest probability within a factor of greater than 2 of the best solver, although it has higher number of wins than either $p = 1.1$ or $p = 2$. Algorithm 4.1 with

$p = 1.1$ and $p = 2$ has comparable performance within a factor of from 2 to 14 of the best solver. Comparing with Fig. 2, we see that for the four p , the lowest probability of Algorithm 4.1 is higher than the highest probability of Algorithm 3.1.

5. Concluding remarks

In this paper, we have extended two effective methods for the MCP (1) studied in [8,9] to a more general case which is based on the generalized Fischer–Burmeister function. We generalize the theoretical results therein, and test the influence of numerical performance by changing the parameter p . The performance profiles of iteration times and function evaluations indicate that the strictly feasible Newton method with $p \in [1.1, 2]$ has better performance, whereas the semismooth Levenberg–Marquardt method with a smaller p , for example, $p = 1.001$, has better performance. Furthermore, comparing Figs. 1 and 2 with Figs. 3–4, we see that the influence of p on the strictly feasible equation-based method is more remarkable than the semismooth Levenberg–Marquardt method. However, different from the merit function method based on ϕ_p (see [19]), there is no evident tendency about the influence of p on the two Newton-type methods.

Acknowledgements

The authors thank the referees for their suggestions about presenting performance profiles which led to much improvement of this paper.

The first author's work is partially supported by National Science Council of Taiwan. The second author's work is supported by National Young Natural Science Foundation (No. 10901058) and Guangdong Natural Science Foundation (No. 9251802902000001).

References

- [1] R.W. Cottle, J.-S. Pang, R.-E. Stone, *The Linear Complementarity Problem*, Academic Press, New York, 1992.
- [2] M. Ferris, J.-S. Pang, Engineering and economic applications of complementarity problems, *SIAM Review* 39 (1997) 669–713.
- [3] M. Ferris, K. Sinapiromsaran, Formulating and solving nonlinear programs as mixed complementarity problems, in: V. Nguyen, J. Strodiot, P. Tossings (Eds.), *Optimization*, in: *Lecture Notes in Economics and Mathematical Systems*, vol. 481, 2000.
- [4] P.T. Harker, J.-S. Pang, Finite dimensional variational inequality and nonlinear complementarity problem: A survey of theory, algorithms and applications, *Mathematical Programming* 48 (1990) 161–220.
- [5] S.C. Billups, S.P. Dirkse, M.C. Soares, A comparison of algorithms for large scale mixed complementarity problems, *Computational Optimization and Applications* 7 (1997) 3–25.
- [6] S.C. Billups, M.C. Soares, QCOMP: A quadratic programming based solver for mixed complementarity problems, *Mathematical Programming* 76 (1997) 533–562.
- [7] M. Ferris, C. Kanzow, T. Munson, Feasible descent algorithms for mixed complementarity problems, *Mathematical Programming* 86 (1999) 475–497.
- [8] C. Kanzow, Strictly feasible equation-based methods for mixed complementarity problems, *Numerische Mathematik* 89 (2001) 135–160.
- [9] C. Kanzow, S. Petra, On a semismooth least squares formulation of mixed complementarity problems with gap reduction, *Optimization Methods and Software* 19 (2004) 507–525.
- [10] C. Kanzow, S. Petra, Projected filter trust region methods for a semismooth least squares formulation of mixed complementarity problems, *Optimization Methods and Software* 22 (2007) 713–735.
- [11] S. Dirkse, M. Ferris, MCPLIB: A collection of nonlinear mixed complementarity problems, *Optimization Methods and Software* 5 (1995) 319–345.
- [12] F. Facchinei, J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vols. I and II, Springer, New York, 2003.
- [13] S. Dirkse, Robust solution of mixed complementarity problems, Ph.D. Thesis, Computer Sciences Department, University of Wisconsin, 1994.
- [14] C. Kanzow, An inexact QP-based method for nonlinear complementarity problems, *Numerische Mathematik* 80 (1998) 557–577.
- [15] O.L. Mangasarian, Global methods for nonlinear complementarity problems, *Mathematics of Operations Research* 21 (1996) 589–614.
- [16] J.-S. Pang, S.A. Gabriel, NE/SQP: A robust algorithm for the nonlinear complementarity problems, *Mathematics Programming* 60 (1993) 295–337.
- [17] J.-S. Chen, The semismooth-related properties of a merit function and a descent method for the nonlinear complementarity problem, *Journal of Global Optimization* 36 (2006) 565–580.
- [18] J.-S. Chen, On some NCP-functions based on the generalized Fischer–Burmeister function, *Asia-Pacific Journal of Operational Research* 24 (2007) 401–420.
- [19] J.-S. Chen, S. Pan, A family of NCP-functions and a descent method for the nonlinear complementarity problem, *Computational Optimization and Applications* 40 (2008) 389–404.
- [20] E.D. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles, *Mathematical Programming* 91 (2002) 201–213.
- [21] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
- [22] L. Qi, *C-differentiability, C-differential operators and generalized Newton methods*, AMR 96/5, Applied Mathematics Report, University of New South Wales, Sydney, 1996.
- [23] J.-S. Pang, L. Qi, Nonsmooth equations: Motivation and algorithms, *SIAM Journal on Optimization* 3 (1993) 443–465.
- [24] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, *Mathematics of Operations Research* 18 (1993) 227–244.
- [25] L. Qi, J. Sun, A nonsmooth version of Newton's method, *Mathematical Programming* 58 (1993) 353–367.
- [26] A. Fischer, Solution of the monotone complementarity problem with locally Lipschitzian functions, *Mathematical Programming* 76 (1997) 513–532.
- [27] C. Kanzow, H. Kleinmichel, A new class of semismooth Newton-type methods for nonlinear complementarity problems, *Computational Optimization and Applications* 11 (1998) 227–251.
- [28] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's method, *SIAM Journal on Numerical Analysis* 23 (1986) 707–716.
- [29] S.C. Billups, Algorithms for complementarity problems and generalized equations, Ph.D. Thesis, Computer Sciences Department, University of Wisconsin, 1995.